Against Intelligence-First

Intelligence as Plastic Feedback and Constraint Navigation (Not Mystical Agency)

T30 · Prime 3 · Structural Clarifier
Jeremy C. Jones (ORCID 0009-0007-2515-3774)
HoldingLight LLC · universalcollapse.com
Version: v1.0 (2026-06) | License: CC BY 4.0
DOI: https://doi.org/10.17605/OSF.IO/RN3TH

Purpose (lens reset). Prevent a recurring interpretive failure: treating "intelligence" as a primitive essence or a synonym for complexity / optimization. Prime 3 gives a disciplined, portable way to talk about intelligence that does not smuggle in consciousness, personhood, or magic. It treats intelligence as a structural capacity that can be present or absent by evidence, not as an inner essence to be intuited.

Abstract

Intelligence is often treated as a private, human-centric possession. This Prime de-privatizes the term by defining it operationally: a system is intelligent when it can update its own constraints using plastic memory and feedback such that it improves prediction / compression, control, and transfer across contexts. On this view, intelligence is not "more coherence" and not "more complexity"; it is coherence that becomes learnable (adaptive constraint management). The Prime supplies (i) a minimal litmus test, (ii) a feedback ladder that separates regulation from learning, and (iii) a compact reporting metric (I*) that turns intelligence claims into something falsifiable and comparable within declared baselines and evaluation suites. It is stated in UCT kernel terms (Ω, K, C^K_t, x_t*, R_t, S_t, T, U).

1. The Recurring Confusion This Prime Corrects

Intelligence-first is not the study of intelligent systems. It is the interpretive move that treats intelligence as a primitive explanation, or as a label earned by complexity alone. This creates predictable category errors:

- **Anthropomorphism:** reading purpose, understanding, or "inner experience" into systems that merely stabilize under fixed constraints.
- **Trivialization:** calling any feedback control "intelligence" (thermostat fallacy) and thereby draining the term of discriminative power.
- **Optimization reification:** mistaking an optimization description (least action, free energy, "the system minimizes X") for an internal learner.
- **Complexity worship:** treating large state space, unpredictability, or emergent patterns as proof of intelligence.
- **Baseline neglect:** claiming intelligence without stating what non-learning baseline is being beaten, and on what transfer suite.

Prime 3 replaces the essence-talk with a structural criterion: intelligence is the presence of an endogenous update loop that modifies constraints in response to evaluative feedback, yielding measurable gains that persist beyond the training context. "Endogenous" qualifies the update mechanism and memory substrate,

not the signal: the evaluative feedback may originate at the boundary or in the environment, but the plastic state change credited to the system must be owned by the system as defined.

What this Prime does not claim

This Prime does not claim that an intelligent system is conscious, that it understands its task, or that it has inner experience, personhood, or rich agency. Nor does it claim that complexity, large state spaces, or optimization-shaped behavior are themselves evidence of intelligence, or that intelligence is uniquely human. It claims something narrower: intelligence is an endogenous plastic update loop that improves performance across contexts — present or absent by structural criteria, measurable against a non-learning baseline, and possible without consciousness or understanding.

2. Intelligence in Kernel Terms

UCT's kernel gives a clean grammar that separates (a) lawful coherence and (b) adaptive intelligence. Given a structured possibility space Ω and active constraint set K, the constraint-conditioned collapse operator selects a realized outcome and writes its records:

$$C^{K}_{t} : \Omega \rightarrow (x_t^*, R_t, S_t, \Omega_{t+1})$$

Here x_t^* is the resolution, R_t the records, S_t the residue, T the record-time, and U the update map that carries outcomes forward into the next constraint state $K_{t+1} = U(K_t, x_t^*, R_t, S_t)$; collapse names the operation (C^{K}_{t}) and resolution its achieved result (x_t^*). Record-time T is not a separate output of the operator but the cumulative record depth: the count or depth of accumulated record layers, not a numerical sum of their contents.

Coherence (Prime 0) is what you get when a system repeatedly collapses under a largely fixed K and thereby stabilizes invariants, attractors, or compressible structure. Intelligence begins when the system itself can modify K (or an internal proxy for K) using feedback — i.e., when U is not merely "the world's update" but becomes an active, plastic mechanism inside the system.

Scale note: intelligence is boundary- and timescale-relative

The system boundary you choose determines what counts as an update loop. If you define the system as a population, selection acting on heritable records is a real update loop. If you define it as an individual organism, intelligence usually refers to within-lifetime learning. If you define it as a deployed model with frozen parameters, do not attribute to it the adaptation that occurred upstream during training. State the boundary before you claim the intelligence.

Scope note: in-context adaptation is not parameter learning

A system can improve within a single episode — using a context window, retrieval store, or temporary memory — while its underlying parameters never change. Separate two layers. Adaptation at the context layer happens inside one episode; adaptation at the parameter layer changes persistent internal state and carries into later episodes. In-context adjustment can clear the first two criteria when the context state lies inside the declared boundary and feedback or task evidence causally shapes it — but it is not parameter learning, and it does not survive the episode boundary: a fresh session begins

> without it. The transfer criterion is what separates the two. So report the layer explicitly: state what adapts (parameter, context window, retrieval store, external memory) and whether the gain survives a cleared context. If it vanishes at reset, call it context-layer adaptation, not cross-episode Prime-3 intelligence.

In plain terms: coherence is constraint made visible; intelligence is constraint made learnable.

3. Minimal Criteria (Litmus Test)

A full Prime-3 intelligence claim requires all three conditions:

- **Plastic memory exists:** internal parameters or state within the declared system boundary can be durably changed relative to the evaluation timescale, and those changes alter future behavior.
- **Feedback or selection drives the update:** an evaluative signal — error, reward, constraint violation, fitness proxy — causally shapes the plastic state. The signal may originate outside the boundary, but the update mechanism being credited must lie inside it.
- **Cross-context improvement is demonstrated:** gains persist on held-out, shifted, or newly encountered tasks or regimes specified before evaluation, rather than merely improving in-place or overfitting the training context. This does not require universal generality — only performance above a pre-specified baseline on conditions not seen during adaptation.

If any condition fails, you may still have strong coherence, sophisticated control, or complex dynamics — but you do not have intelligence in the Prime-3 sense.

4. The Feedback Ladder: From Coherence to Intelligence

To prevent the thermostat fallacy, separate feedback into levels:

- **F0 — Dynamical coupling (no corrective feedback):** the system evolves under K; any "memory" is just state persistence.
- **F1 — Fixed feedback (regulation without learning):** feedback exists, but the policy / parameters are fixed (no plastic update).
- **F2 — Adaptive closed-loop (learning feedback):** feedback drives durable state, parameter, or policy updates that improve performance beyond any fixed policy.

Intelligence begins at F2, and only when paired with plastic memory. F1 systems can be remarkably stable and useful, but they are not learning. F2 does not require consciousness; it requires durable, feedback-driven update.

5. A Prime-Level Metric: I* as an Audit Score

When you want to speak quantitatively, use a baseline-relative score rather than vibe-based claims. Define a non-adaptive baseline B and a test suite E. Then score the adaptive system S by three gains, each defined relative to B:

- C_g **(compression / prediction gain):** does S reduce description length or predictive loss (better modeling of regularities in Ω)?

- U_g **(control / utility gain):** does S achieve task goals more reliably or efficiently (better constraint navigation)?
- T_g **(transfer gain):** do those gains persist under context shift (new regimes, tasks, or distributions)?

Combine them as a weighted sum:

$$I^* = w_c \cdot C_g + w_u \cdot U_g + w_t \cdot T_g \qquad (w_c + w_u + w_t = 1)$$

I^* is domain- and baseline-relative. Each gain should be normalized against the baseline B on the stated evaluation suite E before weighting, and weights should be declared in advance for comparative claims; I^* is comparable across systems only when B, E, normalization, and weights are comparable. $I^* > 0$ means the adaptive loop delivers measurable advantage beyond the best fixed alternative — but on its own it supports an adaptive-gain claim, not a full Prime-3 intelligence claim: that additionally requires nonnegative or positive transfer under the stated transfer regime.

Prime boundary: I^* is not a universal IQ. It is a reporting handle that forces clarity about baselines, contexts, and what "better" means.

Negative transfer. Report negative transfer explicitly; a system that improves in one regime while degrading in others may have $U_g > 0$ but $T_g \leq 0$.

6. Category Errors to Avoid

These are the common interpretive failures that block coherence-first reasoning about intelligence:

- **Optimization description ≠ internal optimizer.** "The system minimizes X" may be an analyst's summary, not a learner inside the system.
- **Coherence ≠ intelligence.** Crystals, convection cells, and stable attractors can be highly coherent with $I^* \approx 0$.
- **Complexity ≠ intelligence.** Huge state spaces, nonlinear dynamics, or unpredictable signals do not imply plastic feedback.
- **Evolution timescale ≠ individual timescale.** A population can exhibit adaptive updating across generations; do not misattribute that update loop to an individual agent or a single episode.
- **Offline training ≠ in-loop adaptation.** A trained model with frozen weights may be competent without being currently intelligent (Prime-3 sense) — unless a specified runtime memory, retrieval store, policy, or other internal state is being updated by feedback inside the declared system boundary.
- **Human-in-the-loop tuning should be reported as such.** If people supply the adaptation, the system itself does not own the intelligence claim.

7. Intelligence Reporting Standard (IRS)

To keep "intelligence" from turning into rhetoric, use a minimum disclosure standard. If you cannot fill this out, downgrade the claim (e.g., "adaptive control present" or "feedback regulation present") instead of declaring intelligence.

- **System and boundary.** State what the system is, what counts as internal state, and what environment / task suite E it is evaluated in.
- **Baseline B.** Specify the best non-adaptive (fixed) baseline you compare against (or S with learning disabled).
- **Plasticity mechanism.** Name what is plastic (weights, gains, policy, genome frequencies, synapses) and how updates persist.
- **Feedback / selection signal.** State the evaluative signal driving updates (error, reward, fitness proxy) and where it enters the loop.
- **Update schedule.** Online or episodic? Continuous or batch? When is learning frozen for transfer tests?
- **Transfer regime(s).** Describe the shift(s) used to assess transfer gain (new tasks, new distributions, new operating regimes).
- **I* report.** Report C_g, U_g, T_g (with uncertainty if possible) and the weights w_c, w_u, w_t.
- **Ablation.** Show that disabling adaptation collapses the gains (or provide an equivalent causal attribution).
- **Resource cost.** Required for comparative claims; otherwise recommended. Report compute time, energy, data, and human intervention needed to achieve the gains (efficiency matters).
- **Leakage / contamination check.** Report whether transfer tasks, user feedback, or test regimes were exposed during training or tuning.

8. How to Use This Prime in the Stack

Prime 3 does not add new law-level claims. It prevents a misread. Cite it when you need to block the move from "complex" to "intelligent" without evidence of plastic feedback; to separate regulation (F1) from learning (F2) in biology, robotics, AI, or cognition discussions; or to justify an I*-style audit in a domain paper without turning that paper into a philosophy debate.

The division of labor across the library: WPs (WP03, WP04) state law-level and domain-level claims and tests; Structural companions (Structural Biology, Structural Mind) provide domain-specific methods and protocols; and Primes provide conceptual hygiene — how not to misread the law or the method. Prime 3 is the lens reset: how not to misread "intelligence" while working under the kernel.

9. One-Line Carryforward

Intelligence is plastic feedback that improves across contexts: coherence made learnable.

On the Prime Series

This is one of five Prime papers. Each clears a single term that minds and frameworks routinely compress in the same way — flattening a structured thing into a primitive and then treating the primitive as bedrock. The five catches are: coherence read as something added rather than constraint made visible (Prime 0); randomness read as irreducible chance rather than a provisional label for unmodeled structure (Prime 1); chaos read as disorder rather than structured unpredictability (Prime 2); intelligence read as essence or

mystical agency rather than adaptive constraint-navigation (Prime 3); and nothingness read as an explanatory ground rather than a structured state mislabeled as absence (Prime 4).

Each Prime is free-standing. It asks no commitment to Universal Collapse Theory, because the cleared term is one the reader already holds and can judge on its own ground. Together, the five form the program's hygiene layer: guardrails against compression, keeping concepts from being mistakenly elevated into primitives — whether by an outside reader meeting the term cold or by a builder working inside UCT. The guardrail serves on both sides of that line: it helps any reasoner about to flatten a term, and it keeps the corpus from drifting on its own vocabulary.

These clarifications are held in the same mode they ask of the reader: provisional, methodological, and open to revision — never final claims about what the cleared term ultimately is.

Appendix A. Micro Examples (Intentionally Small)

These examples only illustrate the litmus test and the reporting posture. They are not a methods manual.

A1. Thermostat vs. self-tuning thermostat

- **Fixed thermostat:** F1 (feedback regulation), but no plastic update. $I^* \approx 0$ relative to the best fixed thermostat.
- **Self-tuning thermostat** (auto-tunes gains based on error statistics): F2 + plastic. Can show $U_g > 0$ under drift (e.g., seasonal changes), and some T_g if it generalizes to new regimes.

A2. Bacterial chemotaxis — adaptive navigation without consciousness

- The bacterium uses sensing to bias motion up a nutrient gradient (feedback signal).
- Plasticity can exist in biochemical adaptation (e.g., receptor methylation states), enabling improved navigation under changing concentrations.
- **Transfer question:** does the bias persist across shifted gradients or environments, beyond what a fixed policy would do? That is where an I^* audit lives.

A3. Frozen model vs. online learner (AI example)

- A deployed model with fixed parameters and no runtime-updated memory or policy can be competent (high coherence in behavior) without current intelligence (no in-loop plasticity).
- An online-learning variant that updates from user feedback or environment reward can qualify (F2) if it beats a frozen baseline and retains gains under transfer tests.
- **Reporting discipline:** always state what is adapting (parameters, prompt memory, retrieval store) and run an ablation (turn learning off).

References (Selected)

Ashby, W. R. (1956). An Introduction to Cybernetics. Chapman & Hall.

Åström, K. J., & Murray, R. M. (2008). Feedback Systems: An Introduction for Scientists and Engineers. Princeton University Press.

Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. Minds and Machines, 17(4), 391–444.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359.

Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction (2nd ed.). MIT Press.

This paper is part of the Universal Collapse Theory library. For a reading guide and full architecture, visit universalcollapse.com/roadmap.

AI Disclosure. AI tools were used to assist with manuscript preparation. The underlying theory, arguments, and interpretive claims are the author's own, and the author takes full responsibility for the manuscript.

Citation: Jones, J. C. (2026). Against Intelligence-First: Intelligence as Plastic Feedback and Constraint Navigation (Not Mystical Agency). Prime 3 (Structural Clarifier). HoldingLight LLC. https://doi.org/10.17605/OSF.IO/RN3TH

Series: Universal Collapse Theory — T30: Ground-Clearing (Prime Papers)